

SozioNet - Web-Ressourcen für ein künftiges Fachportal Sozialwissenschaften

Schmiede, Rudi; Schumann, Natascha; Meier, Wolfgang; Heise, Sue

Preprint / Preprint

Konferenzbeitrag / conference paper

Empfohlene Zitierung / Suggested Citation:

Schmiede, R., Schumann, N., Meier, W., & Heise, S. (2003). *SozioNet - Web-Ressourcen für ein künftiges Fachportal Sozialwissenschaften*. Darmstadt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-255757>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

SozioNet -

Web-Ressourcen für ein künftiges Fachportal Sozialwissenschaften

Natascha Schumann, Wolfgang Meier, Sue Heise, Rudi Schmiede

Institut für Soziologie
Technische Universität Darmstadt
Schumann@ifs.tu-darmstadt.de

Abstract: SozioNet is part of a forthcoming social science information portal, which is currently being developed by the infoconnex initiative. SozioNet provides access to freely available resources with relevance to social sciences. SozioNet aims at building a network of social science institutions and scientists, to agree on and establish common metadata standards. SozioNet implements a general infrastructure for the creation of semantically rich metadata and for harvesting and retrieving relevant resources in a domain specific focus.

1 Einleitung

SozioNet ist ein Bestandteil des im Rahmen von infoconnex entstehenden Fachportals Sozialwissenschaften. SozioNet versteht sich als zusätzliches Angebot zu bereits bestehenden Informationsdienstleistungen im sozialwissenschaftlichen Bereich, wie z.B. den Fachdatenbanken und weiteren Angeboten des IZ Sozialwissenschaften und der Virtuellen Fachbibliothek Sozialwissenschaften ViBSoz.

Das Ziel von SozioNet besteht in der Vernetzung frei zugänglicher Internet-Ressourcen. Die Idee dazu besteht bereits seit 1998; seitdem haben eine Reihe vorbereitender Treffen stattgefunden. Seit April 2002 wird das Projekt SozioNet vom Bundesministerium für Forschung und Bildung im Rahmen des Verbundprojekts infoconnex gefördert. Bislang besteht in den Sozialwissenschaften keine lebendige und etablierte Preprintkultur wie beispielsweise in den Naturwissenschaften. Jedoch wird das Internet auch in diesem Bereich zunehmend dazu genutzt, um unterschiedliche sozialwissenschaftliche Ressourcen zugänglich zu machen. Diese liegen dann auf den Servern der jeweiligen Person oder Institution, so dass sie für Außenstehende nur schwer auffindbar sind. Denn die herkömmlichen Suchmaschinen gehen in der Regel nach algorithmischen Methoden vor und vernachlässigen den fachlichen Kontext. Daher sind wissenschaftlich relevante Quellen auf diesem Wege nur schwer zu finden. Die Auszeichnung von Ressourcen mit standardisierten Metadaten ermöglicht ein Retrieval, das gezielt nach sozialwissenschaftlichen Internetquellen sucht. Inzwischen stellen viele Einrichtungen die unterschiedlichsten Arten von Publikationen und Materialien zur Verfügung: Arbeitsberichte, Lehrmaterialien, Präsentationen, Artikel, Aufsätze, Projektbeschreibungen, Personen- und Institutionenseiten u.v.m.

Das Ziel von SozioNet ist es, einen bequemen Zugang zu Ressourcen zu schaffen, die verteilt auf den Servern der jeweiligen Einrichtungen vorhanden sind. Darüber hinaus möchte SozioNet einen Beitrag dazu leisten, die Sichtbarkeit sozialwissenschaftlicher Einrichtungen und ihrer Aktivitäten zu verbessern. Außerdem sollen Sozialwissenschaftlerinnen und Sozialwissenschaftler motiviert und darin unterstützt werden, ihre Publikationen über das Internet zu verbreiten und mit standardisierten Metadaten auszuzeichnen.

Vorbilder für SozioNet stammen aus der Mathematik (MathNet), aus der Physik (PhysNet) und aus den Sozialwissenschaften (SOSIG - Social Science Information Gateway).

Bislang beteiligen sich zwölf sozialwissenschaftliche Einrichtungen an SozioNet. Sie werden ihre Ressourcen mit Metadaten auszeichnen, Schnittstellen zu bereits in Datenbanksystemen vorgehaltenen Daten herstellen und die von uns bereitgestellten Metadatenschemata und Werkzeuge implementieren, evaluieren und weiterentwickeln.

Langfristig wird eine Etablierung von SozioNet innerhalb der Sozialwissenschaften angestrebt. Dafür ist es notwendig, mit den beteiligten Institutionen Organisationsstrukturen zu schaffen, um eine nachhaltige Nutzung und Erweiterung des Dienstes über die Projektlaufzeit hinaus zu gewährleisten. Hierzu bestehen Vereinbarungen mit dem Informationszentrum Sozialwissenschaften.

2 Zielsetzungen

Das Ziel von SozioNet ist die Vernetzung frei zugänglicher Ressourcen von Wissenschaftlerinnen und Wissenschaftlern sowie von wissenschaftlichen Einrichtungen aus dem Bereich der Sozialwissenschaften im Internet.

Über herkömmliche Suchmaschinen sind solche Ressourcen nur schwer zu finden. Daher sollen diese mit geeigneten Metadaten ausgezeichnet werden, um eine gezielte Suche zu ermöglichen. Mit Hilfe von Metadaten kann bei der Suche der fachliche Kontext berücksichtigt und somit die Relevanz einer Ressource einfacher und schneller beurteilt werden. Ein aufwendiges Sortieren in endlosen Trefferlisten wird so umgangen.

Außerdem sollen interessierte Sozialwissenschaftlerinnen und Sozialwissenschaftler dazu animiert werden, ihre Publikationen mit Metadaten auszuzeichnen und verstärkt über das Netz zu vertreiben. Es soll innerhalb der sozialwissenschaftlichen Community eine überregionale Informationsinfrastruktur geschaffen werden, die von den Wissenschaftlerinnen und Wissenschaftlern selbst gestaltet und organisiert wird. Die Ressourcen selbst verbleiben in der Verantwortung und Pflege der veröffentlichenden Einrichtung oder Person, SozioNet organisiert lediglich den Zugang zu den Informationen und stellt Werkzeuge für die Metadatenerzeugung zur Verfügung.

Eine Voraussetzung für ein erfolgreiches Retrieval im wissenschaftlichen Kontext ist die Anreicherung der jeweiligen Ressource mit zusätzlichen Informationen, also mit Metadaten. Diese ermöglichen einen bequemen und schnellen Zugriff auf verstreute sozialwissenschaftliche Fachinformationen.

Daher ist es notwendig, sich auf einen gemeinsamen Metadatenstandard zu einigen, den alle benutzen und der möglichst viele für den Kontext von SozioNet relevante Aspekte berücksichtigt. Darüber hinaus sollen die verwendeten Metadaten internationalen Standards genügen, um eine Nutzung unabhängig von SozioNet und die Anschlussfähigkeit an zukünftige Entwicklungen zu gewährleisten.

3 Metadaten

Innerhalb von SozioNet kommen Metadaten in zwei Bereichen zur Anwendung. Zum einen werden die vorhandenen Online-Ressourcen mit Metadaten ausgezeichnet, und zum anderen enthält die SozioNet-Seite Metadaten.

Für die Auszeichnung von sozialwissenschaftlich relevanten Online-Ressourcen wurde für SozioNet ein eigenes Metadatenschema (auch Ontologie genannt) entwickelt. Das Metadatenschema beschreibt die verschiedenen Arten von Ressourcen, z.B. „research paper“, „presentation“ etc., mit unterschiedlichen Eigenschaften.

Verwendet werden dafür in erster Linie Dublin Core Metadaten und ihre Qualifier. Als etablierter Standard dient Dublin Core der Erschließung von Internet-Ressourcen. Allerdings werden durch Dublin Core nicht alle fachspezifischen Aspekte abgedeckt; daher werden für die inhaltliche Erschließung zusätzlich der Thesaurus und die Klassifikation des IZ Sozialwissenschaften herangezogen. Diese liegen jeweils in RDF vor.

RDF bedeutet Resource Description Framework und ist eine Beschreibungssprache für Metadaten. Es bietet die Möglichkeit der modularen Definition von Metadaten, d.h. es können bereits bestehende und eigene Schemata eingebettet bzw. modelliert werden. RDF Schema bietet jedoch nur begrenzte Ausdrucksmöglichkeiten in Bezug auf Klassenbildung und Beziehungen. Daher wird als zusätzliche Abstraktionsschicht DAML (DARPA Markup Language) verwendet. DAML bietet zusätzliche Modellierungsmöglichkeiten und ist etwas einfacher in der Handhabung. Die Klassifikation und der Thesaurus Sozialwissenschaften sind daher zur Zeit in DAML verfasst. Da das W3C in absehbarer Zukunft DAML durch OWL (Web Ontology Language) ersetzen wird, werden die in SozioNet verwendeten DAML-Notationen dementsprechend an OWL angepasst werden.

Ein Beispieldatensatz macht diese verschiedenen Bestandteile des Metadatenschemas sichtbar:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:sn="http://www.sozionet.org/1.0/">

  <sn:ResearchPaper rdf:about="http://www.zeitschriftarbeit.de/docs/2-2000/wolf.PDF">
    <dc:title>Das Netzwerk als Signatur der Epoche</dc:title>
    <dcq:alternative>Anmerkungen zu einigen neueren Beiträgen zur soziologischen Gegenwartsdiagnose</dcq:alternative>
    <dc:creator>Wolf, Harald</dc:creator>
    <dc:subject rdf:resource="http://sozionet.org/1.0/topics#10220"/>
    <dc:subject rdf:resource="http://sozionet.org/1.0/topics#1080404"/>
    <dc:subject>Netzwerkgesellschaft</dc:subject>
    <dc:subject>Netzwerke</dc:subject>
    <dc:subject>Netzwerkforschung</dc:subject>
    <dc:subject>Organisationssoziologie</dc:subject>
    <dc:subject>Soziologische Theorie</dc:subject>
    <dcq:abstract>Der Aufsatz beinhaltet Aspekte der Netzwerkdebatte, die für eine soziologische Deutung der Gegenwartsgesellschaft empirisch wie theoretisch wichtig sind....</dcq:abstract>
    <dc:publisher>Landesinstitut Sozialforschungsstelle Dortmund</dc:publisher>
    <dcq:created>1999-07-14</dcq:created>
    <dcq:modified>2000</dcq:modified>
    <dcq:IMT>text/pdf</dcq:IMT>
```

```

<dc:language>de</dc:language>
<dcq:isPartOf rdf:resource="http://www.zeitschriftarbeit.de/">
<dcq:isFormatOf>Zeitschrift Arbeit, Heft 2, Jg.9 (2000), S.95-104
</dcq:isFormatOf>
</sn:ResearchPaper>
</rdf:RDF>

```

4 SozioNet-Seite

Neben den Online-Ressourcen sollen auch institutionelle Informationen über SozioNet recherchierbar sein. Die meisten Einrichtungen halten solche Informationen vor, allerdings ist es sehr schwer, sich einen Überblick über die verschiedenen Institute hinweg zu verschaffen. Außerdem sind diese Informationen in jeder Einrichtung unterschiedlich aufbereitet und an unterschiedlichen Orten zu finden, so dass es für externe Besucher der Seite schwer ist, an die gewünschten Informationen zu gelangen.

Angelehnt an das Konzept der „secondary homepage“ oder auch „MathNet-Seite“ aus MathNet soll die SozioNet-Seite eine Navigationshilfe für die Nutzerinnen und Nutzer sein, um gezielt zu institutsbezogenen Informationen zu gelangen. Die SozioNet-Seite ist in Struktur und Layout einheitlich für jedes Institut. Die Nutzerinnen und Nutzer können so ohne Kenntnisse der Strukturen der jeweiligen Institutsserver problemlos nach den gewünschten Informationen recherchieren.

Zunächst wird diese Seite Links zu den Seiten mit der gewünschten Information enthalten, z.B. über die Mitarbeiterinnen und Mitarbeiter. Sukzessive soll das Metadatenschema mit Objektklassen zu dieser Seite erweitert werden, so dass ein semantisch reichhaltiges Netz von relevanten Informationen entsteht. So wird ein Überblick über die Aktivitäten innerhalb der Sozialwissenschaften ermöglicht, der einen echten Mehrwert für die Nutzerinnen und Nutzer erbringt.

5 Harvesting

SozioNet basiert auf der Selbstorganisation der Wissenschaftlerinnen und Wissenschaftler. Das bedeutet, dass die Ressourcen auf den jeweiligen Institutsservern verbleiben und die Verantwortung und Pflege weiterhin den Autorinnen und Autoren bzw. ihren Institutionen obliegt. Daher müssen die Metadaten der jeweiligen Dokumente von einem Harvester eingesammelt werden.

Das ursprüngliche Harvesting-Konzept bietet nicht die optimale Unterstützung für SozioNet, unter anderem deshalb, weil die Metadaten im SOIF-Format ausgegeben werden. Bei der Konvertierung der Metadaten von RDF nach SOIF besteht die Gefahr eines gewissen Informationsverlustes.

Aus diesem Grunde wurde eine gemischte Lösung entwickelt. Bei Einrichtungen, die ihre Materialien in internen Content-Management-Systemen vorhalten, werden die Metadaten über das OAI-MHP (Open Archives Initiative Metadata Harvesting Protocol) geholt. Die OAI wurde 1999 mit dem Ziel gegründet, wissenschaftliche Publikationen möglichst schnell anderen Wissenschaftlerinnen und Wissenschaftlern zugänglich zu machen. OAI-MHP ist ein anwendungsunabhängiges, standardisiertes Protokoll zum Austausch von Metadaten. Die Einrichtung wird dadurch zu einem Dataprovider und SozioNet zu einem Serviceprovider, der die Metadaten abfragt. In diesem Fall muss der SozioNet-Harvester dann nur noch die Volltextindexierung übernehmen.

Institute, die die Metadaten erst erstellt haben, geben die Adressen der zu durchsuchenden Seiten an, und die Harvesting-Komponente von SozioNet durchläuft diese dann in regelmäßigen Abständen.

Alle eingesammelten Metadaten werden in einer zentralen Datenbank gespeichert. Der Harvester übernimmt die Volltextindexierung und muss auf Updates reagieren. URLs können auch direkt in die Datenbank eingegeben werden, so dass der Harvester nicht vollständig die Webadressen indexiert, sondern nur solche, bei denen bekanntermaßen für SozioNet relevante Ressourcen vorhanden sind.

Da alle Komponenten in SozioNet in XML beschrieben sind, werden die Metadaten-sätze in einer XML-Datenbank vorgehalten.

Die folgende Abbildung gibt einen schematischen Überblick über die Architektur von SozioNet:

SozioNet

Technik: Überblick

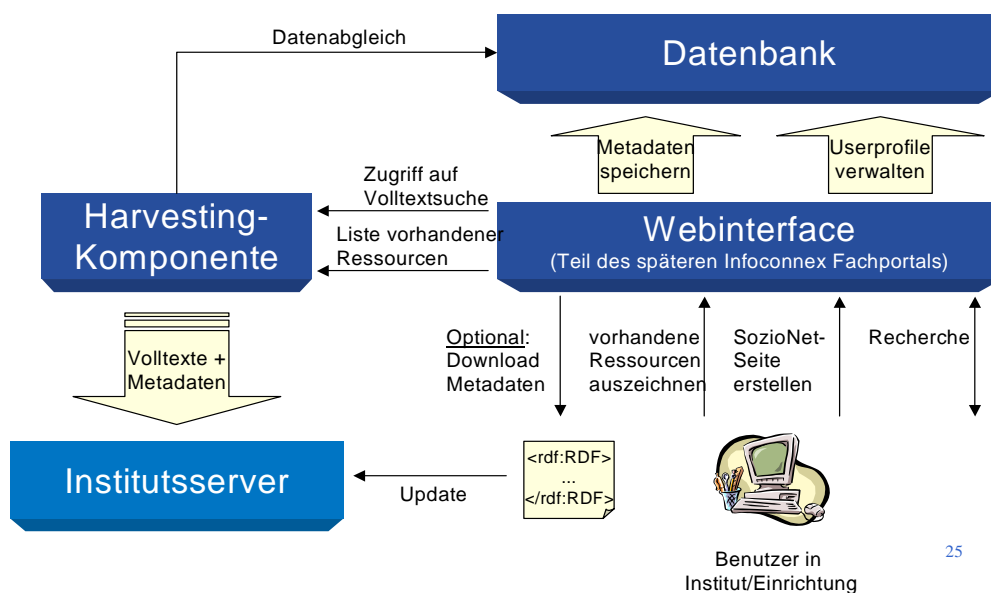


Abb.1: Architektur von SozioNet

6 Aufgaben der Projektpartner

Derzeit beteiligen sich zwölf Einrichtungen an SozioNet. Sie werden ihre Ressourcen so aufbereiten, dass diese über SozioNet recherchierbar sein werden. Außerdem werden sie sich an der Weiterentwicklung des Metadatenschemas beteiligen und die bereitgestellten Werkzeuge evaluieren.

Bei den Projektpartnern handelt es sich sowohl um universitäre Institute als auch um außeruniversitäre Forschungseinrichtungen. Einige Einrichtungen, wie beispielsweise das Deutsche Jugendinstitut in München oder das Wissenschaftszentrum Berlin für Sozialforschung, verwalten ihre Materialien in eigenen Content-Management-

Systemen, während andere erst damit beginnen, ihre Dokumente im Netz zu veröffentlichen.

Dementsprechend ergeben sich für die Projektpartner unterschiedliche Aufgaben. Einrichtungen mit eigenen Systemen müssen entsprechende Schnittstellen zur Verfügung stellen, damit SozioNet auf die vorhandenen (Meta-)Daten zugreifen kann. Es bietet sich an, hier das Open Archives Initiative (OAI) Protocol zu verwenden. Über ein standardisiertes Protokoll (OAI-MHP) können verteilte Metadaten abgefragt und eingesammelt werden. Außerdem müssen Schnittstellen zu den Erschließungswerkzeugen (Klassifikation und Thesaurus) des IZ Sozialwissenschaften entwickelt werden.

Anders sieht die Aufgabenstellung aus, wenn Ressourcen nicht gezielt verwaltet und noch nicht mit Metadaten ausgezeichnet worden sind. In diesem Fall müssen die entsprechenden Dokumente mit Metadaten gemäß der SozioNet Ontologie versehen werden.

Da in der Regel bei Sozialwissenschaftlerinnen und Sozialwissenschaftlern Kenntnisse über Metadatenformate nicht vorausgesetzt werden können, haben wir ein Formular als Eingabehilfe entwickelt. Für die Auszeichnung von Online-Ressourcen mit Metadaten steht der MetaWizard zur Verfügung. In diesem webbasierten Formular spiegelt sich das zugrunde liegende Metadatenchema wider. Die Institute können entweder eigenständig einen dem Metadatenchema von SozioNet entsprechenden Metadatenatz erstellen, oder sie geben die entsprechenden Daten in den MetaWizard ein. Bereits vorhandene Metadaten wie beispielsweise HTML-Metatags werden automatisch extrahiert und in das Formular übernommen. Diese müssen dann nicht mehr zusätzlich eingegeben werden, sondern möglicherweise nur noch angepasst werden.

Neben der formalen Erschließung sollen vor allem auch inhaltliche Merkmale beschrieben werden. Das geschieht zum einen durch die Vergabe von frei zu wählenden Schlagwörtern und zum anderen über die entsprechenden Einträge aus dem kontrollierten Vokabular des Thesaurus und der Klassifikation Sozialwissenschaften. Darüber hinaus sollte ebenfalls ein kurzes Abstract die wesentlichen Inhalte der Ressourcen beschreiben.

Nachdem alle Angaben gemacht worden sind, wird automatisch ein korrekter Metadatenatz generiert. Da jede Nutzerin und jeder Nutzer einen eigenen Account hat, können jederzeit Veränderungen an bereits eingegeben Daten vorgenommen werden. Ebenso kann der Eingabevorgang unterbrochen und ohne inhaltliche Verluste zu einem späteren Zeitpunkt wieder aufgenommen werden.

Der Metadatenatz wird nicht in das Dokument selbst eingebettet, sondern als eigene Datei im selben Verzeichnis gespeichert.

Seite 2: Allgemeine Metadaten

Titel Der Titel des Dokuments (max. 75 Zeichen)	<input type="text"/>
Abstract Kurze Beschreibung des Inhalts des Dokuments	<input type="text" value="SozioNet ist Teil eines im Rahmen von Infoconnexentstehenden Fachportals Sozialwissenschaften, welches die vorhandenen und zukünftigen Informationsdienstleistungen des Faches bündeln wird. Ergänzend zu den bereits bestehenden Informationsangeboten"/>
Autor(en) Format: Nachname, Vorname(n). Zur Eingabe mehrerer Autoren bitte auf das "+"-Symbol klicken.	<input type="text"/>
Herausgeber Die fuer die Veröffentlichung des Dokuments verantwortliche Person oder Institution	<input type="text"/>
Freie Schlagworte Ein oder mehrere frei gewählte Schlagworte zur Beschreibung des Dokuments. Bitte jedes Schlagwort in einem gesonderten Feld eingeben. Um weitere Felder zu erhalten, klicken Sie bitte auf das "+"-Symbol.	<input type="text"/>
Entnommen aus Ist das Dokument die elektronische Kopie eines Originals, tragen Sie hier bitte Angaben zum Original ein	<input type="text"/>
Erstellungsdatum Erstellungsdatum des Dokuments. Bitte Datumsangaben immer in der Form TT.MM.JJJJ oder nur das Jahr (z.B. 2001) eintragen, falls der genaue Zeitpunkt nicht feststeht.	<input type="text" value="06.03.03"/>
Datum der letzten Änderung Wann wurde das Dokument zuletzt geändert? Zum Format der Datumsangabe siehe oben.	<input type="text" value="01.01.70"/>
Sprache des Dokuments	<input type="text" value="Deutsch"/>
Medienformat In welchem Format liegt das Dokument vor (z.B. PDF oder Word)?	<input type="text" value="HTML"/>

Zurück Abbruch Weiter

Abb.2: MetaWizard: Eingabeformular zur Erzeugung von Metadaten

7 Zusammenfassung und Ausblick

SozioNet bietet ein zusätzliches Informationsangebot zu den bereits bestehenden Diensten innerhalb der Sozialwissenschaften an. Es bietet einen schnellen und bequemen Zugriff auf kostenfreie und verteilte Ressourcen des Faches. Die Verwendung von weltweit anerkannten Standards ermöglicht eine nachhaltige Nutzung und gewährleistet, dass die über SozioNet erfassten Daten auch in Zukunft nutzbar sind.

Die Grundkonzeption ist inzwischen abgeschlossen, und die Implementierungsphase hat begonnen. Der MetaWizard kann von den beteiligten Institutionen zur Erzeugung von standardisierten Metadaten gemäß der SozioNet-Ontologie genutzt werden. Dann werden im Verlauf des Jahres 2003 die ersten Ressourcen über SozioNet recherchierbar sein.

Die Struktur für die SozioNet-Seite ist ebenfalls festgelegt und wird den beteiligten Einrichtungen zur Verfügung gestellt, so dass man sich schnell einen Überblick über die verschiedenen Institutionen und ihre Angebote in Forschung und Lehre verschaffen kann.

Weiterführende Links

SozioNet

<http://www.sozionet.org>

infoconnex

<http://www.infoconnex.de>

ViBSoz

<http://www.vibsoz.de>

IZ Sozialwissenschaften

<http://www.gesis.org/IZ/>

MathNet

<http://www.math-net.de>

PhysNet

<http://physnet.net>

SOSIG

<http://www.sosig.ac.uk>

Dublin Core

<http://dublincore.org>

W3C

<http://www.w3.org>

OAI

<http://www.openarchives.org>